IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Carlos Vinicius
2022-02-06

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Discussion
- Conclusion
- References
- Appendix

Executive Summary

- We gathered data from SpaceX API and webscrapped from Wikipedia
- Objective was to analyze Falcon 9 data on good and bad landings

* Analysis revealed that since 2013, landing successes keeps increasing

* Orbit and launch site also have significance in determining landing success

Introduction

- We want to do data analysis and prediction analytics using SpaceX data, to determine what factors contribute to landing successes.

* We would use those factors to propose a new approach for company SpaceY

- What are the launch sites with worse landing success rates? Why?
- What orbits are most likely to success?
- How to pick a machine learning model to predict the outcome of a landing?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX API
  - Wikipedia
- Perform data wrangling
  - Treated null values in PayloadMass column by replace with mean
  - Create column Class with values 1(landing success) or 0(landing failure)
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Use hyperparameter tunning with GridSearchCV method on four different classifiers: SVM, Logistic Regression, Decision Tree and KNN

Data Collection

- We collected data using SpaceX API from https://api.spacexdata.com/v4/ with requests Python Library
- We scrapped data from https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Data Collection – SpaceX API

- requests.get("https://api.spacexdata.com/v4/rockets/"+str(x)).json()
- requests.get("https://api.spacexdata.com/v4/launchpads/"+str(x)).json()
- requests.get("https://api.spacexdata.com/v4/payloads/"+load).json()
- requests.get("https://api.spacexdata.com/v4/cores/"+core['core']).json()
- https://github.com/viniciuscva/capstone_project_winning_space_race_with_data_science/blob/master/Data%20Collection%20-%20API.ipynb

Data Collection - Scraping

- BeautifulSoup Python library
- static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"
- requests.get(static_url)
- BeautifulSoup(content.content, 'html5lib')
- soup.find_all('table',"wikitable plainrowheaders collapsible")
- https://github.com/viniciuscva/capstone_project_winning_space_race_with_data_science/blob/master/Data%20Collection%20-%20Web%20Scraping.ipynb

Data Wrangling

- Generate column Class from landing_outcomes, as values 1(success) and 0(failure)
- Evaluate success rate and values frequency in columns
- https://github.com/viniciuscva/capstone_project_winning_space_race_with_data_science/blob/main/Data%20Wrangling.ipynb

EDA with Data Visualization

- Scatter plots, Line plots, bar charts
- Show the relationship between variables in dataset
- Do feature engineering
- https://github.com/viniciuscva/capstone_project_winning_space_race_with_data_science/blob/main/EDA%20with%20Visualization.ipynb

EDA with SQL

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first successful landing outcome in ground pad was acheived.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order
- https://github.com/viniciuscva/capstone_project_winning_space_race_with_data_science/blob/main/EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- Added markers, marker clusters, circles, lines, coordinates, etc to Folium Map, in order to enrich geospatial data visualization.
- https://github.com/viniciuscva/capstone_project_winning_space_race_with_data_science/blob/main/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

Build a Dashboard with Plotly Dash

- Dashboards with the main plots for visualizing the contributing factors to landing outcome, and its respective magnitude.
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

- Split the dataset into training, cross-validation and testing sets
- Used GridSearchCV for hyperparameter tunning
- Tested four classifiers: Logistic Regression, SVM, KNN and decision trees
- https://github.com/viniciuscva/capstone_project_winning_space_race_with_data_science/blob/main/Predictive%20Analysis.ipynb

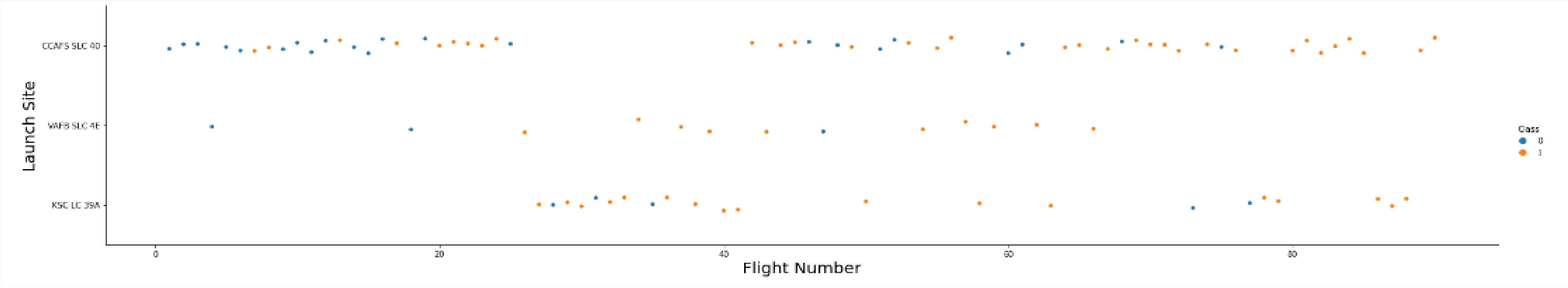
Results

- Exploratory data analysis shows that Launch Site, Payload mass, Orbit and Date have influence on landing outomes
- Success rates keeps increasing since 2013
- Orbits ES-L1,GEO,HEO and SSO have highest success rates
- Decision tree classifier got the best accuracy on test set (94.44%)

Section 2

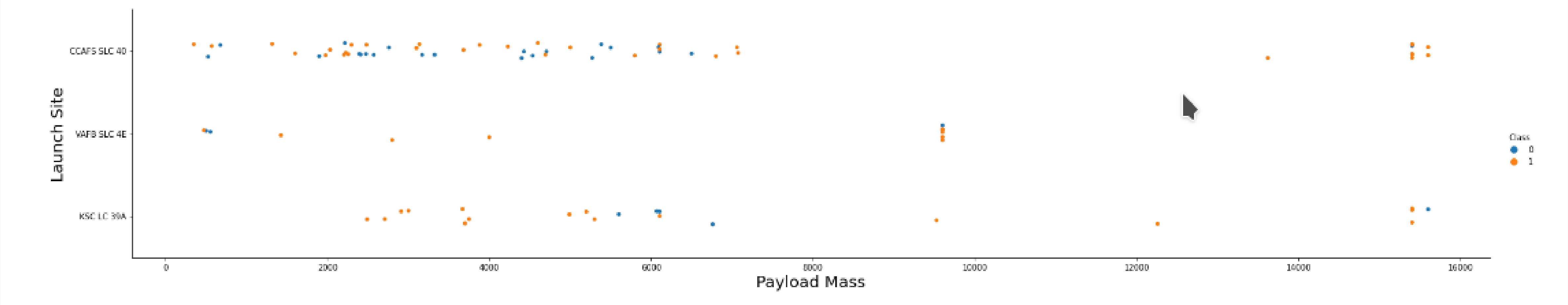
Insights drawn from EDA

Flight Number vs. Launch Site



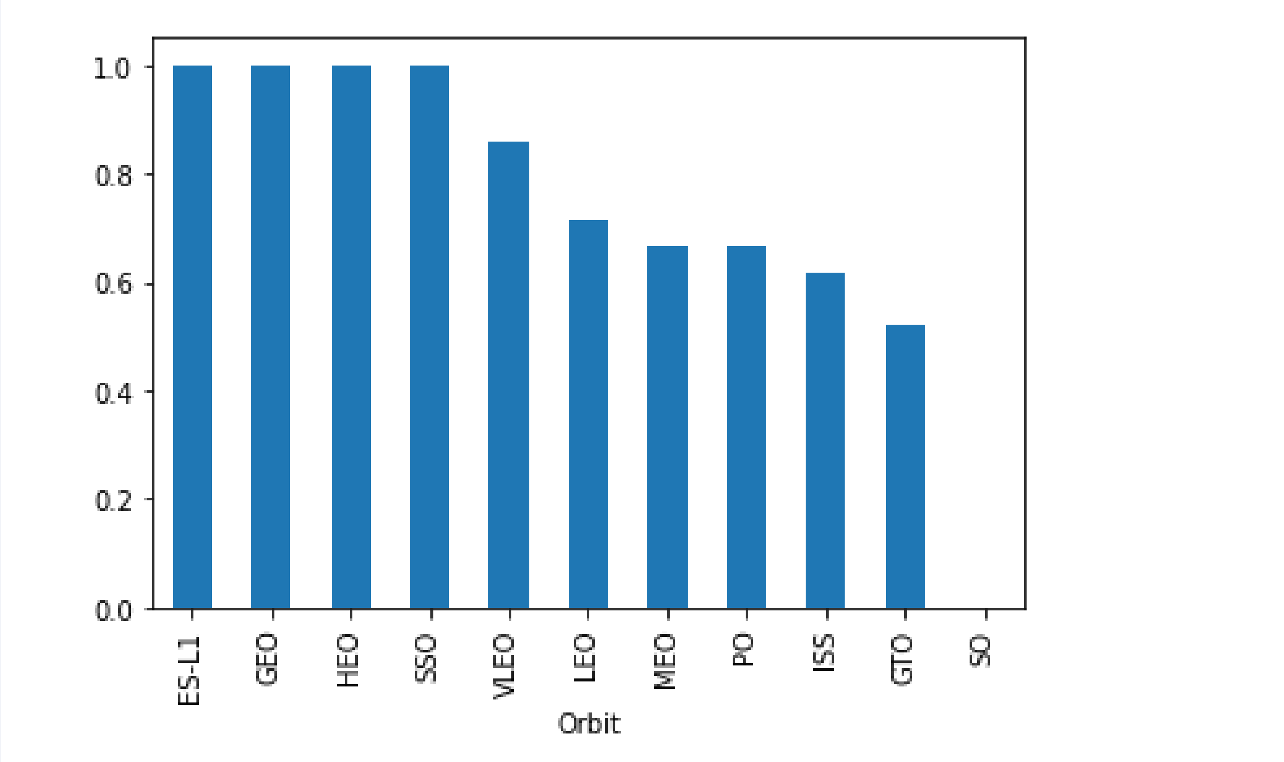
- CCAFS SLC 40 has the most flights

Payload vs. Launch Site



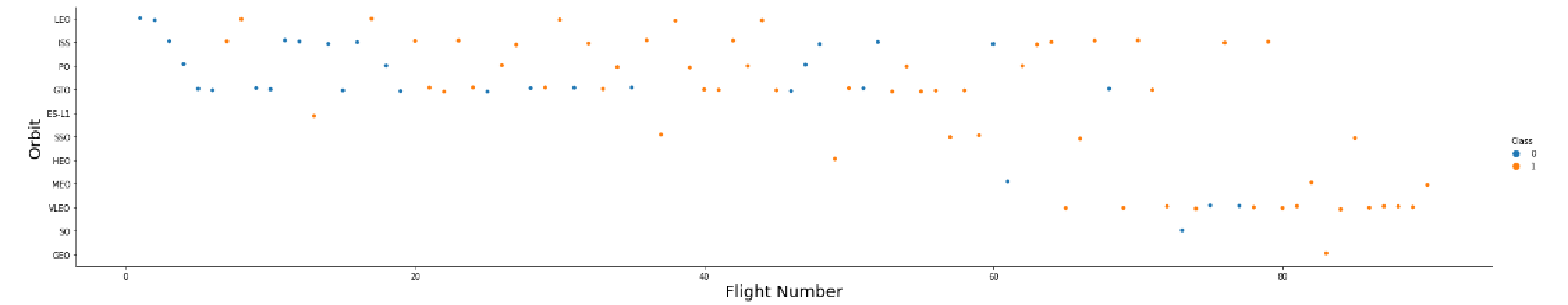
- for the VAFB-SLC launchsite there are no rockets launched for heavy payload mass(greater than 10000).

Success Rate vs. Orbit Type



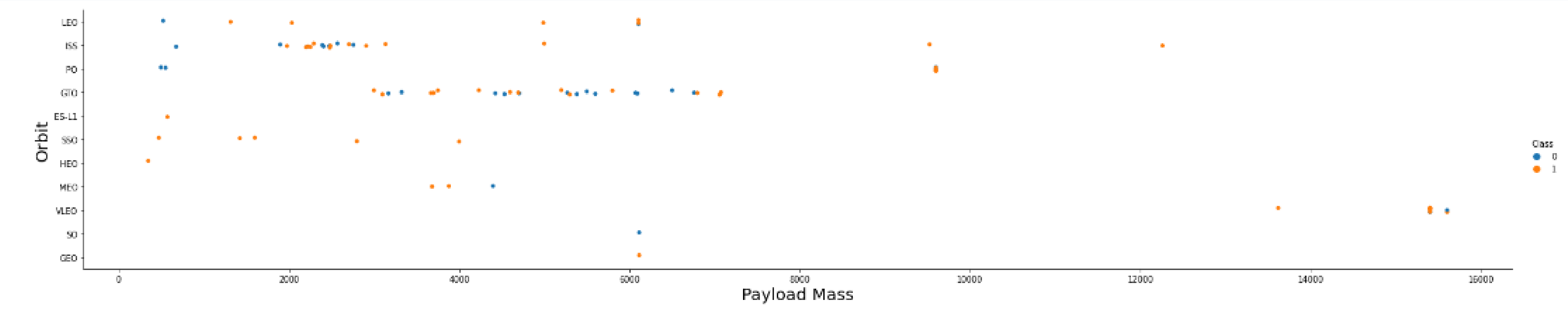
- Orbits ES-L1, GEO, HEO and SSO have the highest success rates

Flight Number vs. Orbit Type



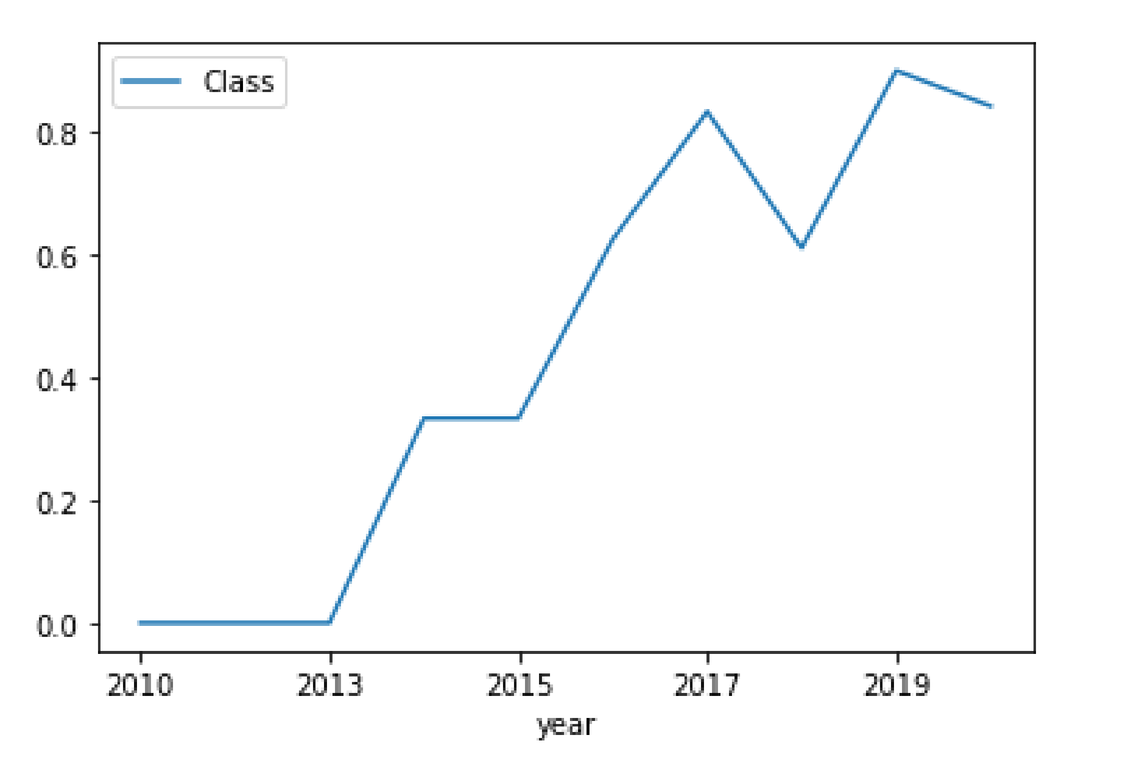
- in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type



- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

Launch Success Yearly Trend



- Success rate keeps increasing since 2013.

All Launch Site Names

```
%%sql
SELECT DISTINCT LAUNCH_SITE FROM SPACEXDATASET
```

```
 * ibm_db_sa://zkm30294:***@0c77d6f2-5da9-48a9-81f
Done.
```

launch_site
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

- All distinct launch sites

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

```
%%sql
SELECT * FROM SPACEXDATASET WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5
```

```
 * ibm_db_sa://zkm30294:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08k
Done.
```

DATE	time__utc_	booster_version	launch_site	payload	paylo
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

```
%%sql
SELECT SUM(PAYLOAD_MASS__KG_) AS TOTAL_PAYLOAD_MASS_FROM_NASA_CRS FROM SPACEXDATASET WHERE CUSTOMER = 'NASA (CRS)'
```

```
 * ibm_db_sa://zkm30294:**[illegible]databases.appdomain.cloud:311
Done.
```

total_payload_mass_from_nasa_crs
45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

```
%%sql
SELECT AVG(PAYLOAD_MASS__KG_) AS AVG_PAYLOAD_MASS_FROM_F9_V11 FROM SPACEXDATASET WHERE BOOSTER_VERSION = 'F9 v1.1'
```

```
 * ibm_db_sa://zkm30294:**[illegible]databases.appdomain.cloud:31198
Done.
```

avg_payload_mass_from_f9_v11
2928

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

```
%%sql
SELECT DATE FROM SPACEXDATASET WHERE UPPER(LANDING__OUTCOME) LIKE '%SUCCESS%' ORDER BY DATE LIMIT 1
```

```
 * ibm_db_sa://zkm30294:**[illegible]databases.appdor
Done.
```

DATE
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

```
%%sql
SELECT BOOSTER_VERSION FROM SPACEXDATASET WHERE LANDING__OUTCOME = 'Success (drone ship)' AND (PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000)
```

```
 * ibm_db_sa://zkm30294:***[illegible]databases.appdomain.cloud:31198/bludb
Done.
```

booster_version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

```
%%sql
SELECT (SELECT COUNT(*) FROM SPACEXDATASET WHERE LANDING__OUTCOME LIKE '%Success%') AS TOTAL_SUCCESSES, (
```

```
 * ibm_db_sa://zkm30294:***[illegible].databases.appdomain
Done.
```

total_successes	total_failures
61	10

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

```
%%sql
SELECT BOOSTER_VERSION FROM SPACEXDATASET WHERE PAYLOAD_MASS__KG_ = (
  SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXDATASET
)
```

```
 * ibm_db_sa://zkm30294:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io
Done.
```

booster_version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

F9 B5 B1049.7

Task 9

List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

```
%%sql
SELECT BOOSTER_VERSION, LAUNCH_SITE, LANDING__OUTCOME FROM SPACEXDATASET WHERE LANDING__OUTCOME = 'Failure (drone ship)' AND YEAR(DATE) = 
```

```
 * ibm_db_sa://zkm30294:***[illegible].databases.appdomain.cloud:31198/bludb
Done.
```

booster_version	launch_site	landing__outcome
F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

```
%%sql
SELECT LANDING__OUTCOME, COUNT(*) AS COUNT FROM SPACEXDATASET WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY LANDING__OUTCOME OI
```

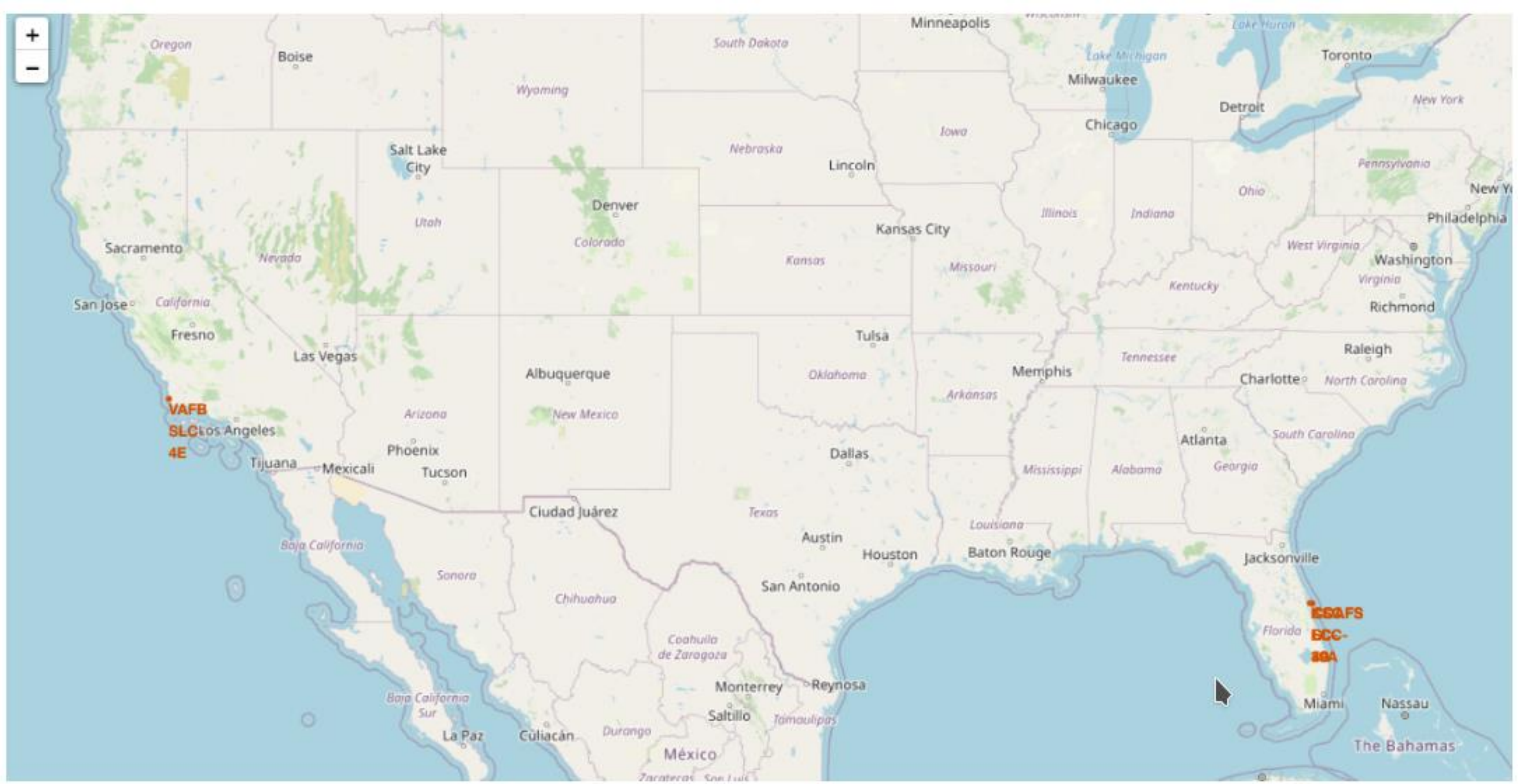
```
 * ibm_db_sa://zkm30294:***[redacted].databases.appdomain.cloud:31198/bludb
Done.
```

landing__outcome	COUNT
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 3

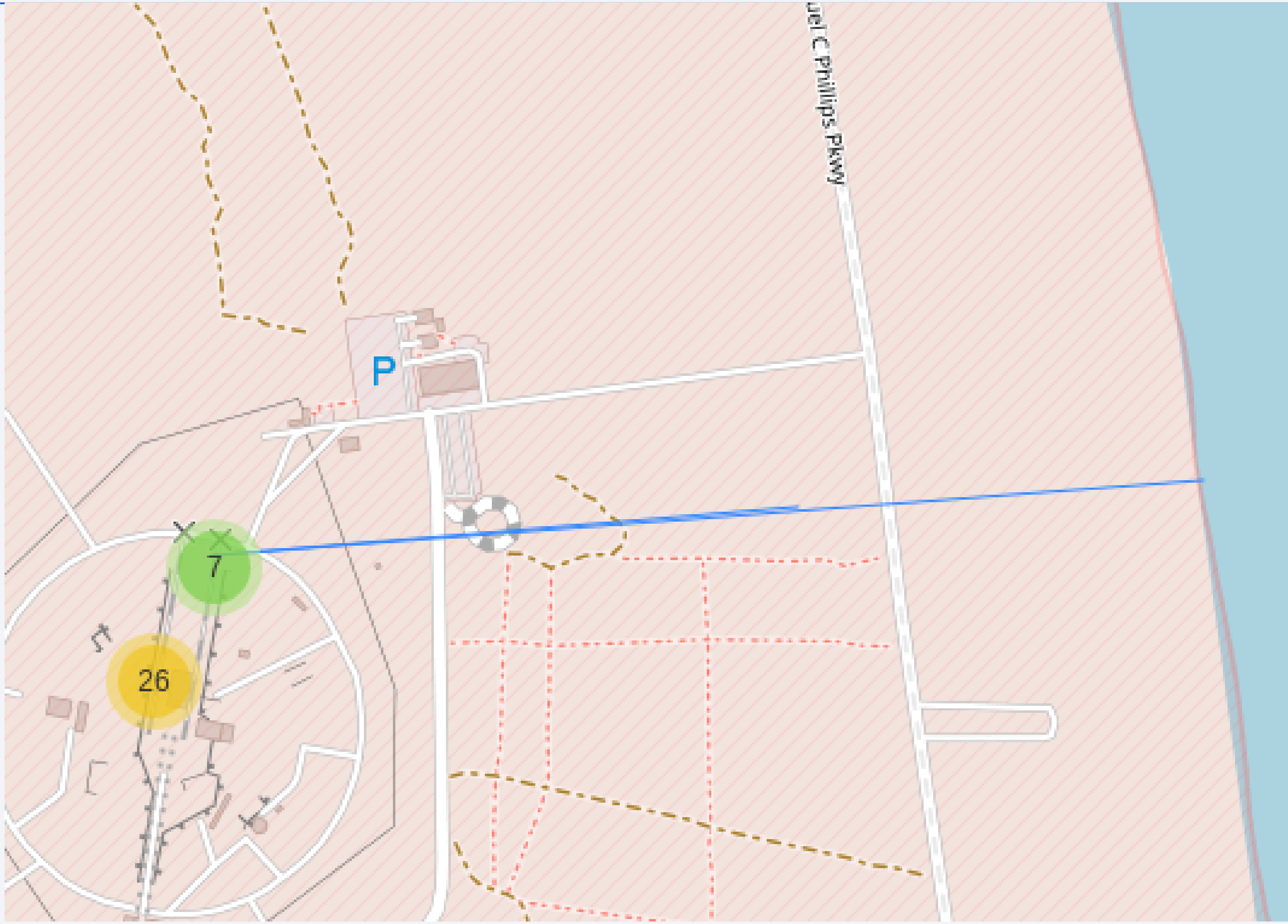
Launch Sites Proximities Analysis

Launch Sites



Markers for landing outomes by launch site

Proximity to coastlines

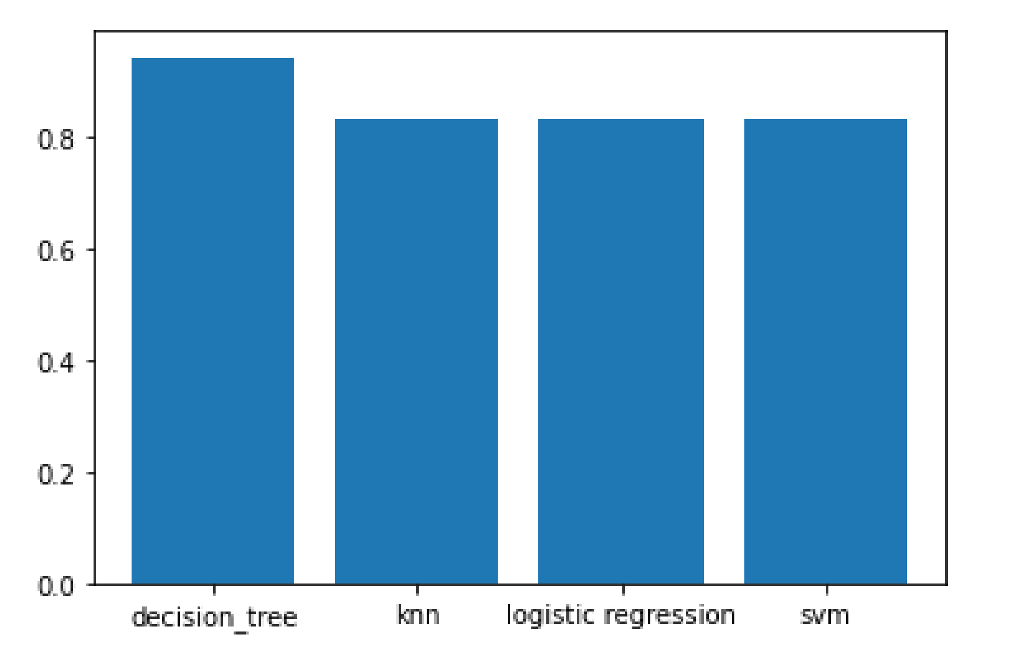


Section 5

Predictive Analysis (Classification)

Classification Accuracy

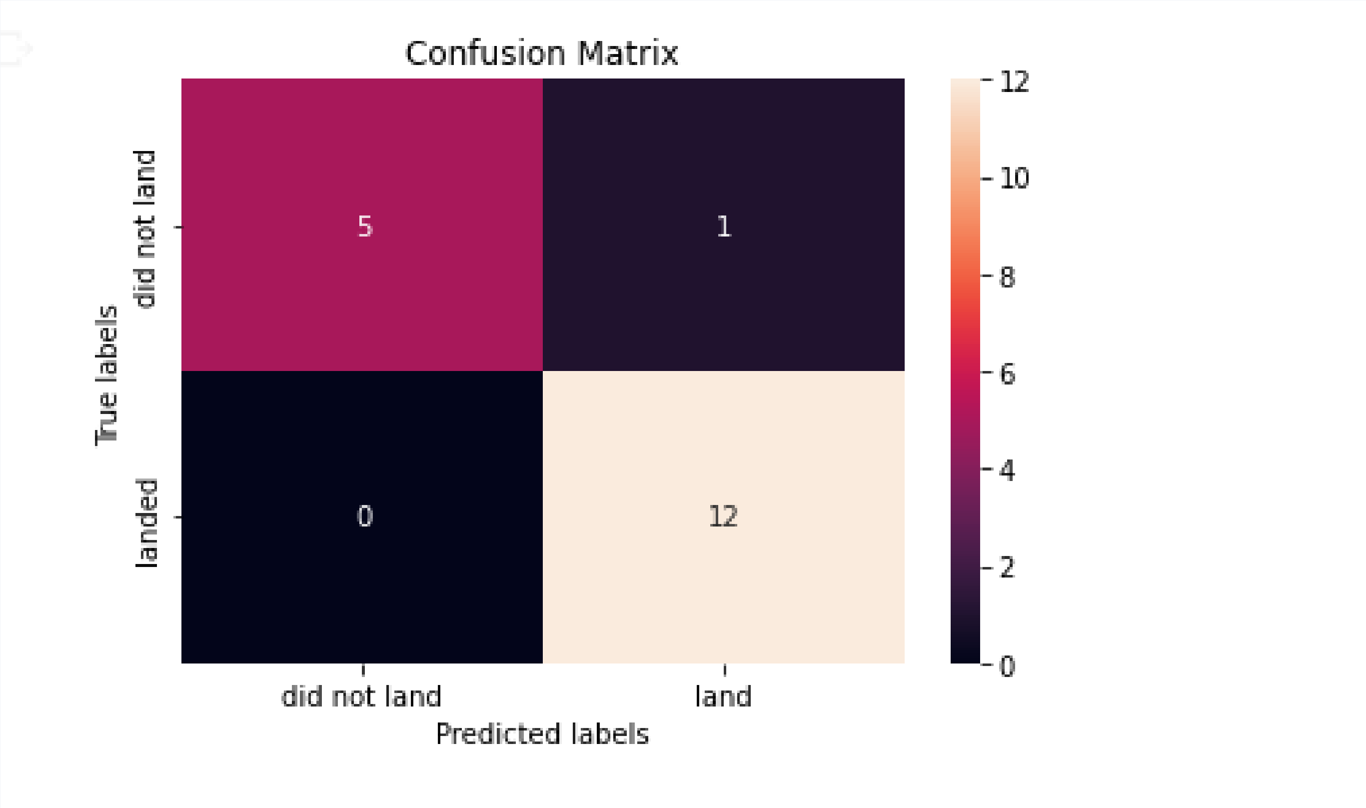
- Visualize the built model accuracy for all built classification models, in a bar chart



- Decision tree classifier has achieved the highest classification accuracy

Confusion Matrix

- Show the confusion matrix of the best performing model (decision tree)



Conclusions

- Decision tree is the best classifier (94% accuracy)
- Exploratory data analysis shows that Launch Site, Payload mass, Orbit and Date have influence on landing outcomes
- Success rates keeps increasing since 2013
- Orbits ES-L1,GEO,HEO and SSO have highest success rates

Appendix

- Link to repository: https://github.com/viniciuscva/capstone_project_winning_space_race_with_data_science

Thank you!